

講義コンテンツの階層構造を用いた参照スライド抽出方式

桐原 牧紀¹ 王 元元² 河合由起子^{3,4} 角谷 和俊¹ ¹関西学院大学 ²山口大学 ³京都産業大学 ⁴大阪大学

研究背景

- e-Learningの普及により授業形式が対面から同時双方向またはオンデマンドの形式に変化
- 柔軟かつ効率的な質問対応が必要
→大量の質問に対して、柔軟かつ効率的な対応(回答)が必要



研究目的

スライドに出現する全キーワードに対する説明にあたる参照スライドをまとめて推薦

学生がスライド10を見ている時、“疑似プログラミング”についての疑問点を持ったがスライド12で“疑似プログラミング”について詳しく書かれている
→しかしスライド10を見ている学生はスライド12はまだみていないので“疑似プログラミング”について完全に理解していないまま講義が進む

Slide 10: アルゴリズム (Algorithm) - 「アルゴリズム」とは... アルゴリズムの表現方法...
Slide 11: アルゴリズムの例(日本語) - 最大値を求める...
Slide 12: アルゴリズムの例(疑似プログラミング言語) - 最大値を求める... (Red box highlights the title)

使用するデータ

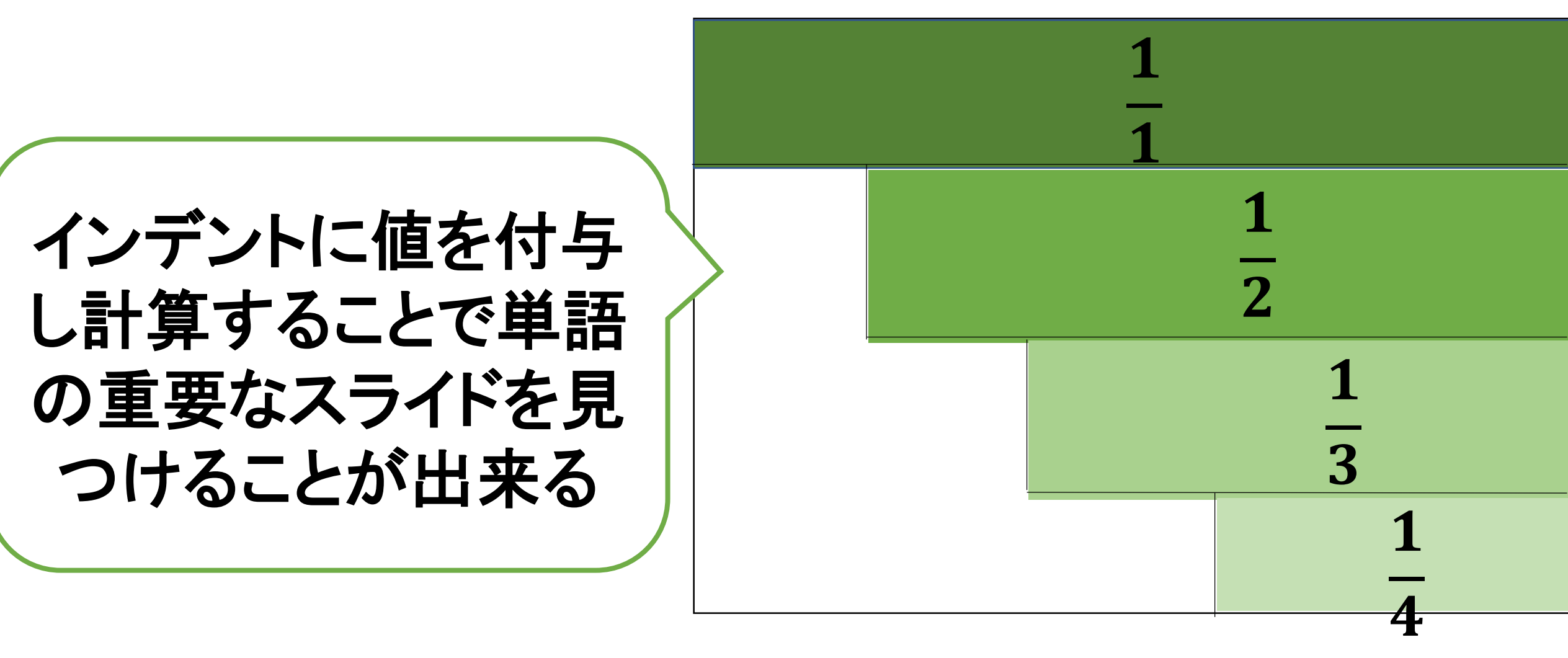
講義スライド中の階層構造テキストデータ

人の頭の中にある処理の手順との違い

- 人が13枚のカードを並び替える (Title: 大きいインデント)
- カードを広げてざっとみて、小さいカードを左に、大きいカードを右に、順番に並ぶように移動させる (Title: 小さいインデント)
- プログラムが13枚のカードを並び替える (Title: 大きいインデント)
- まず、13枚のうちの1枚を取得する (Title: 小さいインデント)
- つぎに、13枚のうちの次の1枚を取得する
- それらと比較して...
- (小さいロジックの積み重ね)

提案手法

- 1枚のスライドでの単語の重要度
(1枚のスライドでのインデントの階層の値の和) ÷ (1枚のスライドでの単語の出現頻度)
- 2枚のスライド間での単語の重要度
(スライドのインデントの値の和) - (スライドのインデントの値)
 $b - a$



①単語Xがタイトルに出てきた後、次のスライドでタイトルと大インデントに出現した場合

a $\frac{1}{1}$ b $\frac{1}{1} + \frac{1}{2} = \frac{3}{2}$

$\frac{3}{2} - \frac{1}{1} = \frac{1}{2}$ $X = \frac{1}{2}$

②単語Xが大きいインデントに出てきた後、次のスライドで大インデントと小さいインデントに出現した場合

a $\frac{1}{2}$ b $\frac{1}{2} + \frac{1}{3} = \frac{2}{3}$

$\frac{2}{3} - \frac{1}{2} = \frac{1}{6}$ $X = \frac{1}{6}$

③単語Xが小さいインデントに出てきた後、次のスライドで小さいインデントと小さい字に出現した場合

a $\frac{1}{1}$ b $\frac{1}{1} + \frac{1}{2} = \frac{3}{2}$

$\frac{3}{2} - \frac{1}{1} = \frac{1}{2}$ $X = \frac{1}{2}$

今後の課題

- 音声での単語の出現頻度を用いた重要度判定を行う
- スライドをある定義づけた意味的まとまりに分け、どのまとまりが最も学生の質問に対して良いものなのかをランキング形式で推薦する
- 2枚のスライド間での単語の重要度応用し、補足スライドを含む意味的まとまりにも対応させ、学生に対して重要なスライドの意味的まとまりを推薦する

謝辞

本研究の一部は、2020年度国立情報学研究所公募型共同研究(20FC04)の助成を受けたものである。ここに記して謝意を表す。